U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 207

A Special Satellite-Radiosonde Colocation Program

Norman A. Phillips
Kenneth A. Campana
Development Division

AUGUST 1979

1. Overall description

This colocation system matches satellite retrievals of layer temperature and precipitable water with a selected set of 20 radiosonde stations. This preselection of radiosondes makes this system different from other colocation systems. Its advantage is that, when combined with a horizontal analysis of the local retrievals, it enables attention to be focused on areas (e.g., oceanic) of special interest to NMC and does so with a better accuracy and a higher rate of colocation than is possible when reasonably small space windows (e.g., 100 km) are demanded from conventional colocation systems. The disadvantages are that retrieval paths (e.g., "clear column," "N*," and microwave) are blended to some extent, and also that it cannot be used to determine horizontal correlations of satellite errors.

It consists of four programs:

Program I

An initialization package to set up station parameters (station numbers, latitude, longitude, etc.) on SYSDB, to set all data records on SYSDB to a value of 7777, and to set various time parameters on SYSDB.

Program II

Run twice daily. Gets radiosonde data from the operational ADP data set, converts it into a satellite-like format, and stores it on SYSDB. It also gets 12 hours of satellite data analyzed to the radiosonde locations and puts it on SYSDB.

Program III

Run approximately every 10 days. Takes 10 days of accumulated radiosonde and satellite data from the SYSDB file and, by time interpolation of the radiosondes, creates matched pairs of satellite-radiosonde layer temperatures and precipitable water values. These pairs are written on a tape, and the SYSDB is restored for the next 10 days of data. Preliminary statistics for each station are printed out, together with a notice of excessive differences between satellite and radiosonde.

Program IV

A statistics package that uses the matched pairs put on tape by Program III to compute detailed statistics by station and by station groups. Editing (deletions) of data pairs corrupted by radiosonde errors that were uncovered by Program III can be performed.

A typical sequence of operations will be

1. Program I
2. Program II 22+ times at approximately 12-hourly intervals
3. Program III
4. Program II 20+ times again
5. Program III

   etc.

Program IV can be run at any time from the tape after step 2 above. Program I can be modified by temporary code changes to alter station numbers in mid-stream (if this will not corrupt the statistics).

Program II uses the subroutine W3FI40 to access satellite data from the NMC/EDIS disk. Subroutine W3AQ06 is used by programs II and III to obtain the current value of time from the computer clock. Program II uses this time value to decide whether the radiosonde data set being sought from the ADP file (copied onto scratch disk by EXEC NWSCOPYS) is old enough (9 hours) to be complete, and if enough time has elapsed (9 hours) for the desired 12 hours of satellite data to have been processed by NESS. A convenient time to routinely run Program II is around 0000 and 1200 GMT ($\pm$2 hours). Each run will then obtain the radiosondes from 10-14 hours earlier, and satellite data from the 12-hour period ending at that radiosonde time. Program II will recover automatically from one missed run (doing twice as much work the next time), but of course cannot recover radiosonde data more than 24 hours old nor old satellite data that has been replaced by NESS with newer data on the NMC/EDIS disk. This recovery feature can be exploited to run Program II in an over-the-counter mode at 18-hour intervals, if desired for convenience. In any event, the only effect of permanently missed data is to eliminate those matched pairs that depend directly on that missed data--later pairs will be unaffected.

The system will handle data from one or from two satellites on NMC/EDIS. CPU times for the programs when listed below are estimated values for two satellites.

2. Initialization Program I

The card deck for this SYSDB initialization job (with code on cards) is as follows:

```
//WWNPTINI JOB (WD23008AC60100A,WWB4-D2),'PHILLIPS',
//             REGION=300K,CLASS=C,TIME=1
//       EXEC NFORXCLG,PARM.FORT='OPTIMIZE(3)'
//FORT.SYSIN  DD *
```

--Code cards for MAIN, and subroutines TIMCAL and CALTIM----

```
/*
//GO.FT30F001  DD DSN=W.NWS.W324.KCRADIO,UNIT=SYSDB,DISP=SHR
//GO.FT40F001  DD DSN=W.NWS.W324.KCSAT,UNIT=SYSDB,DISP=SHR
//GO.SYSIN  DD *
```

---21 data cards----

```
/*
//
```

The subroutine TIMCAL converts a single integer value of time, expressed in seconds from 0000 GCT 1 Jan 1979, into the six integers: year of century, month, day, hour, minute, second. CALTIM performs the opposite conversion. They will work through 1998.

The first data card (format 6I5) is an exact 12-hourly date-time group (e.g.: 80, 2, 28, 12, 0, 0 for the year, month, day, hour, minute, and second). The first raobs collected by Program II will be 12 hours later than this, and the first 12 hours of collected satellite data will begin at the time on this data card.

The remaining 20 data cards (format 4I10) define the 20 selected radiosonde stations. The four integers for each card are:

1. Latitude in hundredths of degrees (- for Southern Hemisphere)

2. Longitude in hundredths of degrees (- for west longitude)

3. Five-digit station number[1]

---

[1]Artificial numbers 99215, 99195, 99211, 99212, 99217 are used for permanent ships P, C, L, M, and K.

4. An integer denoting

>0: Only retrievals over land will be used in colocations for this station

0: Only over-water retrievals will be used for this station

<0: Both land and water retrievals will be used for this station.

Program II assumes that the first five stations are ships, whose reported position on any one observation may differ somewhat from the conventional standard position defined for it in the input data cards from Program I.

The SYSDB area consists of two data sets preset by the two Fortran direct access I/O statements (cf. the GO.FT JCL cards)

```
DEFINE FILE 30(820,60,L,INAP1)

DEFINE FILE 40(880,60,L,INAP2)
```

The 60-byte blocks are read, written, and interpreted as "records" of length 30 I*2 words.

The first set (KCRADIO) contains, in terms of successive 30 I*2 word records, two types of information.

1. The first 20 records, stored in core as IREC(30,20) contain reference data and control parameters initially supplied from the 21 data cards by Program I, and updated by Program II as it is run. It carries control continuity between successive runs of Program II.

2. The remaining 800 60-byte blocks of KCRADIO are used to accumulate radiosonde data. Each ascent is compressed by Program II into 30 I*2 words modeled after the format used by NESS for satellite retrievals on the NMC/EDIS disk. Each radiosonde station 1-20 is stored in sequence: blocks 21-40 containing the first observational set of the 20 stations, blocks 41-60 the second set (12 hours later), up to a maximum of 40 successive 12-hourly collections. Unavailable stations remain on the disk as 30 words, 7777 (i.e., missing), in their prescribed location.

The second data set (KCSAT) on SYSDB is used by Program II to accumulate successive satellite reports in the 30 I*2 word format, in the order in which they are processed. (These are synthesized satellite data analyzed horizontally to the radiosonde location.) The first 80 blocks on KCSAT were originally set up to accumulate running statistics by Program III. This possibility is not being used. The remaining 800 blocks are used for the satellite data, and are capable of accumulating an average of 40 reports per station.

Integer time values are expressed in all codes in two forms:

a. t = seconds since 0000 GCT 1 Jan 1979. When necessary, this is stored as two I*2 words, t_1 and t_2: $t = t_2 + 32768*t_1$.

b. τ = number of 12-hourly intervals since 0000 GCT 1 Jan 1979.

τ is used primarily for radiosonde time identification, the possibility of radiosondes being observed at slightly earlier or later than normal times being ignored. Neither t_1, t_2, nor τ will exceed 32768 during the 20th century.

When the first 20 blocks of KCRADIO are read into the I*2 core array IREC(30,20), this array contains the following information:

1. IREC(1,1) and IREC(2,1): t_1 and t_2 for the last radiosonde set "captured" on KCRADIO.

2. IREC(3,1) and IREC(4,1): Ditto for satellite data (referring to the beginning of the 12-hour interval over which satellite data is collected).

3. IREC(5,1): Block or record number to be used to put the next satellite record on KCSAT (initially 81).

4. IREC(6,1): Number of matched satellite-radiosonde station collections that have been put on the current tape by Program III (initially zero).

5. IREC(7,1): τ value for the radiosonde report for station 1 that goes into block 21 of KCRAD

6. IREC(8,1) = 12345 if system is "active"

   $\neq$ 12345 if Program II should immediately abort when run.

   (This feature allows one to abort an operationally scheduled automatically running job by using an appropriately changed version of Program I to store a number other than 12345 in this location on SYSDB.)

7. IREC(ℓ,2) and IREC(ℓ,3), ℓ = 1,...,20:
   Five digit station numbers in the 32768 $n_1 + n_2$ format.

8. IREC(ℓ,4) ℓ = 1,...,20: The land-ocean tag (+1,0, or -1) for each station.

9. IREC(ℓ,17) and IREC(ℓ,18), ℓ = 1,...,30: A compressed I*2 copy of the last three reported latitudes and longitudes of each of the five ships, and the associated time.

10. IREC(ℓ,19), ℓ = 1,...,20: Latitude in hundredths of degrees for each station (normal value for the five ships.

11. IREC(ℓ,20), ℓ = 1,...,20: Longitude in hundredths of degrees for each station (normal value for the five ships).

All non-used positions in IREC are zero. All eleven types of information are initialized by Program I. Program II updates items 1, 2, 3, and 9. Program III updates items 1, 2, 3, 4, and 5. Program IV does not change any items.

3. Program II

CPU time for this program varies between 8 and 9 seconds. The JCL for this program using load modules is as follows:

```
//WWKCTIRN JOB (WD23008AC60100A,WWB4-D2),'CAMPANA',
//            REGION=230K,CLASS=C,TIME=1
//*FORMAT PR,DDNAME=,DEST=WWB3729
//       EXEC NWSCOPYS,D='NWS.NMC.PROD.ADPUPA.T12Z.LATEST',T='&&ADP12'
//       EXEC NWSCOPYS,D='NWS.NMC.PROD.ADPUPA.T00Z.LATEST',T='&&ADP00'
//        EXC NFORXLG
//LKED.SYSLIB  DD
//             DD
//             DD DSN=NWS.NMC.W3LIB.LOAD,DISP=SHR
//LKED.KCLIB  DD DSN=W.NWS.W324.NPFCST.LOAD.DISP=SHR
//LKED.SYSIN  DD *
//ENTRY MAIN
 INCLUDE KCLIB(NPCOLOC)
/*
//GO.FT10F001  DD DSN=&&ADP00,DISP=SHR
//GO.FT20F001  DD DSN=&&ADP12,DISP=SHR
//GO.FT30F001  DD DSN=W.NWS.W324.KCRADIO,DISP=SHR
//GO.FT40F001  DD DSN=W.NWS.W324.KCSAT,DISP=SHR
//GO.FT88F001  DD  DISP=SHR,DSN=NSS.PSATAT.TOVS.NMCHF      DIRECTORY
//GO.FT89F001  DD  DISP=SHR,DSN=NSS.PSATAT.TOVS.NMCEDS         DATA
//GO.SYSIN  DD *
    1    2
    .6
    .05
   5.
 100.
/*
//
```

NPCOLOC consists of a main code and six subroutines:

TIMCAL
CALTIM
RADIO
SATCOL
ANAL
ITSOL

The five data cards are

1. (2I5) Allowed satellite identification numbers.

2. (F10.5) SCAN = radius in thsds of kms of circle about each radiosonde station within which satellite retrievals will be collected for analysis.

3. (F10.5) WINDOW = distance from the station in thsds of km such that if at least one retrieval is at least this close to the radiosonde location, the requirement for a retrieval in each of four quadrants around the radiosonde will be ignored (see text near the end of section 3).

4. (F10.5) ANAT

5. (F10.5) ANAW

These last two are the critical values (in 0.1 degrees and millimeters precipitable water) used by SATCOL to test the rms expected analysis error that ANAL returns to SATCOL along with each value ANAL has analyzed to the radiosonde location. The colocation for that level is cancelled if these values are exceeded.

The main code of Program II has the following functions:

1. Print out current time from computer clock.

2. Read in IREC from KCRADIO on SYSDB and reassemble into a format in COMMON that is more readily usable by subroutines RADIO and SATCOL.

3. Stop if IREC (8,1) is not equal to 12345.

4. Read in the five data cards.

5. Decide if the older of the two ADP radiosonde collections has already been processed. If not, call RADIO to process it.

6. Repeat 5 for the newer of the two ADP sets, except that it will not be processed until it is at least 9 hours old.

7. Update IREC (1,1) and IREC (2,1) to record any processing of radiosondes accomplished in 5 or 6 above. Also update the ship locations in IREC(ℓ,17) and IREC(ℓ,18). Copy the updated IREC onto KCRADIO.

8. Call SATCOL to collect successive 12 hour groups of satellite data which have not yet been processed and which are at least 9 hours old. Update (IREC (3,1), IREC (4,1)) and IREC (5,1), and write IREC on KCRADIO, after each call to SATCOL.

9. Stop. This is normally done with a "STOP 7777" Fortran statement, but after 22 radiosonde collections have been put on KCRADIO, a simple "STOP" is used. In a 360/195 job net, this feature allows automatic calling of Program III with its tape every 10 days without asking for unnecessary twice-daily mounting of the tape used in that program.

Subroutines RADIO and SATCOL write radiosonde and satellite data on KCRADIO and KCSAT in 60-byte blocks consisting of 30 I*2 words. (The satellite data written on KCSAT has been analyzed to the radiosonde location.) The first 23 words of each block carry the following information:

Word	Radiosonde and Satellite
1	Station number (1-20)
2	Station latitude (.01 degrees)
3	Station longitude (.01 degrees)
4	t_1
5	t_2

Observation time = $32768\ t_1 + t_2$

Layer mean virtual temperatures (in units of 0.1°K):

6	1000-850 mb
7	850-700
8	700-500
9	500-400
10	400-300
11	300-200
12	200-100
13	100-70
14	70-50
15	50-30
16	30-10
17	10-5
18	5-2
19	2-1
20	1-0.4

	Layer Precipitable water (mm)
21	Surface-700 mb
22	700-500
23	500-300

The remainder of the radiosonde record has total precipitable water (mm) in word 24, surface temperature (.1°C) in word 25, and surface pressure in 0.1 mb in word 30. Words 26-29 are 7777.

Words 24-27 are 7777 in the satellite record. Word 28 for the satellite data is given a value e, that provides information on the retrieval method used for the tropospheric (p > 100 mb) temperatures. The original satellite retrievals are assigned e values as follows,

100: microwave + HIRS'

200: N*

300: clear column

These e values are analyzed to the radiosonde location also. The single analyzed e value appearing in the 30-word record constructed by SATCOL will therefore have a value ranging from 100 to 300, and will suggest the mix of retrieval methods that has gone into that composite retrieval. It will be used in Program IV to stratify statistics according to retrieval method.

Word 29 is the estimated surface pressure assumed by NESS from the (known) orography at the retrieval location. Word 30 is an integer equal to

$$\text{satellite identifier} + 16*n_t + 256\ n_s$$

where n_t and n_s are the number of retrievals that have gone into the reconstructed analyzed retrieval for T at the layers 200-100 mb (characteristic of the troposphere) and 1-0.4 mb (characteristic of the stratosphere).

Subroutine SATCOL operates as follows:

1. Given the initial time t_o in seconds of a desired 12-hourly collection interval, an outer loop uses W3FI40 to search for retrievals within (SCAN x 1000) km of each of the 20 stations. This search is done three times; first for a short period Δt of about 200 seconds preceding t_o, then for the 12 hours following t_o, and finally for the period of Δt seconds after (t_o + 12 hours). The first and third of these intervals allow retrievals to be captured as an entire group from an orbit that passes a station very close to 0000 or 1200 GMT.

The locations used for the five ships in the first and third search are those reported for the radiosondes at the beginning and end, respectively, of the central 12 hour collection period. For the central 12 hour search, the average location of the ships at the two end points is used. If a ship radiosonde report was missed, the standard location is used.

2. A retrieval in these time ranges is saved, up to a maximum of 40 per station, if

   a. it is within SCAN*1000 kms of a station

   b. it comes from an acceptable satellite number

   c. its land/ocean tag in retrieval word 8 fits the land/ocean criteria (-1,0,+1) for the station in question (given on original data cards).

Retrievals are saved in the condensed 30 I*2 format described earlier.

3. After completion of 1 and 2, the chronological order of the retrievals collected for a particular station is determined. Up to a maximum of four co-location events are then determined for each station. This is done by noting points in the chronological sequence at which the retrieval time increases by an amount somewhat greater than twice the time required for the satellite to traverse a distance equal to the collection radius, SCAN. This results, for a given station, in an ordering of its nearby retrievals into 1,2,3, or 4 batches, each batch consisting of data taken from one orbit of one satellite near the station. (A batch is dropped if all of its retrievals lie just outside the center 12-hour period, in order to avoid duplication of colocations from the previous or next run of Program II.)

4. Temperature data for the bottom standard layer is included in the original individual satellite retrievals by SATCOL only if the NESS reported surface pressure for that sounding is sufficiently close to 1000, 850, 700 mb, etc.

5. Each observed variable in words 6-23 and 28-29 of the 30 I*2 word retrieval records in one batch is then analyzed by subroutine ANAL to the station location. The single retrieval constructed from this batch is made up of these analyzed values. The associated time value (in words 4 and 5) is the average t of the retrievals in the batch.

6. The horizontal coordinates x and y with respect to the radiosonde station for each retrieval in a batch presented to ANAL for analysis is expressed by SATCOL in thousands of kilometers, using an approximation to a local stereographic cartesian projection centered at the station:

$$x = 6.371\,\lambda\,(\cos\theta_0 - s\,\sin\theta_0)$$

$$y = 6.371\,(s + 1/2\,\cos\theta_0 \sin\theta_0\,\lambda^2)$$

in which

θ_0 = station latitude

s, λ = difference in latitude and longitude (radian) between satellite report and station.

7. SATCOL prints out for each station a 30-word copy of the individual retrievals saved in step 2 above, and a copy of the final reconstructed retrievals valid at the station locations. These, together with the individual radiosonde records printed out by RADIO, must be referred to in judging whether any large radiosonde-satellite discrepancies uncovered by Program III are due to erroneous radiosonde data and therefore should be ignored (edited out) by Program IV.

Subroutine ANAL is given the data, level by level, from a satellite station batch (expressed as deviations from the mean value), together with the eastward (x) and northward (y) coordinates of each data point from this station. An analyzed value is returned only if

a. There is a report in all four quadrants (considering quadrants oriented according to the x,y axes, or, if this fails, using axes at 45°)

b. Failing a, at least one point is closer than the distance set by WINDOW (data card).

The analysis to the origin of the deviation is based on optimum interpolation principles (assuming no error), using a simple Gaussian correlation function.

$$r_{ij} = \exp(-2d^2)$$

where d is the distance between points i and j in thousands of kilometers. The subroutine ITSOL developed by K. Bergman is used to invert the correlation matrix. It returns a normalized estimate of the interpolation error, which, when unnormalized by multiplication with the standard deviation of the data, is used by SATCOL as a test. If this value is larger than that set by the data cards ANAT and ANAW, the analyzed value is discarded by SATCOL.

The correlation function given above was found to reproduce quite well the results from several subjective hand analyses of satellite data.

Subroutine RADIO:

This subroutine searches the appropriate ADP file (00Z or 12Z) for each of the twenty radiosonde stations. Both mandatory and significant level data are extracted from the station report (see NMC Office Note 29) and changed to a satellite sounding format. The radiosonde "raw" data, including NMC quality control flags, is printed for each station. The final "pseudo-satellite" radiosonde report is also printed, before writing it to disk (FT40F001).

Geopotential heights at mandatory pressure levels are used to obtain thickness temperatures for as many of the previously described "satellite" layers as possible. Reported temperatures and dewpoint depressions are used to construct precipitable water in three layers. NMC quality control flags are used to determine the "goodness" of the data.

Some important details of RADIO are:

1. Permanent ships which are off location by more than 1° latitude or longitude are classified as moving ships by NMC. In searching for the five permanent ships, RADIO will look at moving ships too. Any permanent ship off location more than 2.5° latitude and (or) 4.0° longitude is considered missing.

2. Geopotential heights are accepted as "good" data when the quality control flags are:

A or I = passed vertical consistency check with tight limits

H = manual HOLD inserted by Aviation Branch

Temperatures are accepted when flagged A, I, or H.

3. The first significant level has no quality flags and is accepted as the earth's surface.

4. Heights and temperatures must have <u>at least</u> one "good" vertically adjacent neighbor (i.e., neither missing nor flagged as bad data), otherwise they are not accepted.

5. <u>Gross Error Check</u>:

Reported temperatures and computed thickness temperatures are flagged as bad data if they are outside the range +100°C to -50°C.

Dewpoint depressions greater than 40 degrees are not accepted. Depressions between 30 and 40 degrees are considered "motorboating" values--a relative humidity of 15% is assumed.

6. Specific humidity is computed at each temperature level and the profile is vertically sorted according to pressure. Precipitable water is then computed in each radiosonde layer. The precipitable water sounding must cover the layers surface-700 mb and 700-500 mb (within 20 mb of the boundary pressure) in order to be saved for disk storage. The 500-300 mb layer must at least have data up to 400 mb--any amount above 400 mb should be quite small for our purpose.

7. Total precipitable water in the column surface-300 mb is saved in word (24) only if all three sub-layers have good data. Surface temperature and pressure are saved in words (25) and (30)--tenths of °C and tenths of mb.

8. A list of the missing radiosonde stations is printed at the completion of RADIO.

4. Program III

The CPU time for this program is about 24 seconds (18 of which are for compilation). Its JCL is as follows:

```
//WWNPT1DT  JOB (WD23008AC60100A,WWB4-D2),'PHILLIPS',
//             REGION=300K,CLASS=C,TIME=1
//       EXEC NFORXCLG,PARM.FORT='OPTIMIZE(3),XL'
//FORT.SYSIN DD DISP=SHR,DSN=W.NWS.W324.KCTIROS.SOURCE(NPTIMCAL)
//           DD DISP=SHR,DSN=W.NWS.W324.KCTIROS.SOURCE(NPCALTIM)
//           DD *
```

---Code cards (main only)----

```
/*
//LKED.SYSLIB  DD
//             DD
//             DD DSN=NWS.NMC.W3LIB.LOAD.DISP=SHR
//GO.FT30F001  DD DSN=W.NWS.W324.KCRADIO,DISP=SHR
//GO.FT40F001  DD DSN=W.NWS.W324.KCSAT,DISP=SHR
//GO.FT50F001  DD DSN=NPTIROS,UNIT=TAPE9,DISP=(MOD,KEEP),
             LABEL=(1,SL,,OUT),VOL=SER=E15699
```

The two JCL cards for the tape (GO.FT50 etc.) shown here are for a second or later transfer of data by this program from disk to tape E15699. The first time Program III operates on a new tape these two JCL cards should be replaced by the three cards:

```
//GO.FT50F001  DD DSN=NPTIROS,UNIT=TAPE9,DISP=(NEW,PASS),
//           LABEL=(1,SL,,OUT),VOL=SER=E15699,
//           DCB=(RECFM=VBS,BLKSIZE=2404)
```

This program has the following functions:

1. Take 10 days of radiosonde and satellite retrievals that have been accumulated on KCRADIO and KCSAT, and, by interpolation in time of the radiosondes, construct radiosonde satellite pairs that are matched in time as well as space.

2. Compare the radiosonde and satellite values, tabulate preliminary individual station statistics, and print out information concerning excessive differences.

3. Write the pairs on a tape, following whatever pairs were written on the tape by the previous run of Program III.

4. Delete the 10 days of processed data on the SYSDB data sets KCSAT and KCRADIO, rearrange any new data, and update IREC (the first 20 blocks of KCRADIO). The next run of Program II will continue where it left off, but with the recently completed 10 days of data removed.

The time interpolation process to match a satellite report at time t is linear , and considers the radiosondes at t_1 and $t_2 = t_1 + 12$ hrs, where $t_1 < t < t_2$. If one of t_1 or t_2 is missing, a time window is allowed within which the radiosonde that is present is accepted as valid. Poleward of 30°, this window is ±1 hour. Equatorward of 30° it is ±6 hours. (Most equatorial stations take only one radiosonde per day.)

Information is written on the tape in sequential records of 30 I*2 words. Each station having at least one pair of matched radiosonde-satellite data will have the following records copied on to the tape.

An identification record. This contains

Word 1 = -9999 (to identify an identification record)

2 = station number (1-20)

3 = number of pairs for this station that follow this identification record

4 & 5 = τ values at beginning and end of 10-day period

6-30 = zero

This is followed by the satellite member (a 30 I*2 record) of the first pair, the radiosonde member of the first pair, the satellite member of the second pair, etc. Stations that have no pairs in this 10-day period do not have anything written to tape.

As the tape is written, the counter NTPREC carried in IREC(6,1) is increased by one each time a collection of pairings for a station is written on the tape.

The code contains an option of being run without writing on tape and without modifying IREC and the data on SYSDB. This option is exercised by setting the test word IEX equal to +1 instead of -1 in the Fortran cards located several places after the 190 CONTINUE Fortran card. This option is convenient for testing modifications to Program III without corrupting the data, or for preliminary looks at the data before the 10 days have elapsed.

The station statistical printout includes a copy of the first 16 words of both members of each matched colocation pair, and the 28th word of the satellite member of the pair (the e value denoting retrieval type). For each of the 15 layer temperatures and three precipitable water values, the number of pairs, the radiosonde mean, the satellite mean, the rms difference, and the max and min differences are printed out.

Furthermore, any temperature difference or water difference in excess of a test criterion is printed out, together with the radiosonde and satellite data involved. (The test criteria, expressed as squares of the satellite units of 0.1 degrees and millimeters of water, are set by two DIFOK = Fortran

statements just after statement 421.) Given the time of the report (especially the τ value), one can then examine the two original radiosonde reports as they were originally printed by Program II, and determine whether the radiosonde report in the matched pair was in error because of an error in one of the original radiosondes. If this is deemed to be the case, editing cards should be added to the data cards for Program IV. (The most frequent problem has been caused by the practice of the Aviation Branch of inserting H (for "hold") at all levels for a radiosonde in the ADP data, overriding the other quality tags inserted by the automatic processing codes. This can occasionally cause problems at those levels that are not examined by the Aviation Branch. Ship PAPA is often given H, but its importance to the satellite statistics does not allow the remedy of treating H's as an unacceptable report by subroutine RADIO in Program II.)

5. Program IV

This program computes stratified statistics from the matched pairs on the tape generated by Program III. Its CPU time is 10 seconds for each 10 days of data processed plus 17 seconds for compilation. Its JCL is as follows:

```
//WWNPTITA JOB (WD23008AC60100A,WWB4-D2),'PHILLIPS',
//            REGION=300K,CLASS=C,TIME=1
//       EXEC NFORXCLG,PARM.FORT='OPTIMIZE(3),XL'
//FORT.SYSIN DD DISP=SHR,DSN=W.NWS.W324.KCTIROS.SOURCE(NPTIMCAL)
//           DD DISP=SHR,DSN=W.NWS.W324.KCTIROS.SOURCE(NPCALTIM)
//           DD *

---Code cards (MAIN only)----

/*
//LKED.SYSLIB  DD
//             DD
//             DD DSN=NWS.NMC.W3LIB.LOAD,DISP=SHR
//GO.FT30F001  DD DSN=W.NWS.W324.KCRADIO,DISP=SHR
//GO.FT40F001  DD DSN=W.NWS.W324.KCSAT,DISP=SHR
//GO.FT50F001  DD DSN=NPTIROS,UNIT=TAPE9,DISP=(OLD,PASS),
//           LABEL=(1,SL,,IN),VOL=SER=E15699
//GO.SYSIN DD *

---Data cards----

/*
//
```

The data cards are as follows:

Cards 1 and 2 (6I5): The date-time group (year of the century, month, day, hour, minute, second) for the beginning and for the end of the period for which statistics are desired.

Card 3 (I10): The number of editing cards that follow (0 if none)

Editing Cards (5I10)

Four station grouping cards (4I5)

Satellite number card (2I5)

Tape record card (I10).

The five integers on an editing card are as follows:

n_1 Station number (1-20)

n_2,n_3 Time of report to be deleted in format $t = 32768*n_2 + n_3$

n_4,n_5 Sequence of "levels" 6-23 to be deleted in the matched 30I*2 pair defined by (n_1, n_2, n_3). For example, the combination $n_4 = 6$ and $n_5 = 6$ will delete only the 1000-850 mb mean temperature. The combination $n_4 = 7$, $n_5 = 8$ will delete both the 850-700 and the 700-500 mb mean temperatures. (Separated nonadjacent levels cannot be selectively deleted, unless all intervening ones are also.)

For each of the four station grouping cards, the program interprets the four integers n_1, n_2, n_3, n_4 on such a card as follows. The sequence nl, $n_1+1,\ldots,n_2$ plus n_3, $n_3+1,\ldots,n_4$ defines a group of stations. For example, 6, 7, 8, 8 identifies stations 6, 7, and 8 as the station group under consideration. The program will compute overall statistics from this station group. If fewer than four station groupings are desired, four cards must still be submitted, but the program will ignore any station grouping card that is identical to its predecessor.

The next data card defines two satellite identification numbers, k_1 and k_2, say. The outermost loop of the program is determined by these. If these two numbers differ, the program will run three times, first accepting only colocation pairs from the tapes that have a satellite number k_1, a second time accepting only pairs from satellite k_2, and a third time accepting pairs from both k_1 and k_2 satellites. If the two numbers are identical, the program will run only once, accepting pairs from only that single satellite defined by $k_1 = k_2$.

The last data card contains an integer that can override the number "NTPREC" otherwise gotten by Program IV from IREC(6,1) on SYSDB. NTPREC is used by Program IV to define the number of batches of station colocation pairs that have been written on the tape that it will use (E15699 in the JCL example shown). However, IREC(6,1) on SYSDB might well apply to a newer tape. If the number on this last data card is L.T.E.zero, the SYSDB value of NTPREC will be used by Program IV. If the card number is G.T.zero, it will be used for NTPREC instead of the SYSDB value. (Note: Program III updates the SYSDB value of IREC(6,1) for the tape it uses, and prints out the updated value.)

Statistics are also printed for each station included in a statistical subgroup. The station heading information for these tables is carried by 20 format statements (numbered 21-40) in the code. These must be updated by hand whenever Program IV is run on a tape having a new station list.

The statistical measures printed out include the number of observations, the mean of the satellite data, the mean of the radiosonde data, the mean difference, the rms difference, the variance of the satellite data about its mean, the variance of the radiosonde data about its mean, and the ratio of the last two (satellite ÷ radiosonde).

For individual stations, the mean value of the retrieval parameter e is printed out. For the multi-station group statistics, the collected statistics for each group are separated into three further subsets according to the e-value associated with each satellite report.

$e > 250$: mostly clear column retrievals = "PATH 1"

$150 \leq e \leq 250$: indeterminate = "PATH 2"

$e < 150$: mostly microwave retrievals = "PATH 3"

Vertical correlation of the satellite-radiosonde temperature differences are also printed out for the station subgroups.

Acknowledgments

A. Desmarais provided advice on the design of the SYSDB-tape system and on its routine operation. J. McDonell provided initial advice on the quality control flags used by the ADP system at NMC, and was also able to respond to several of the oversights we uncovered in that procedure. K. Bergman provided his subroutine ITSOL.